



Capítulo 10

Análisis Discriminante

Capítulo 10

Análisis Discriminante

1. Introducción

El **Análisis Discriminante** es una técnica estadística multivariante que busca una doble finalidad:

- En primer lugar, el análisis discriminante es una técnica, sustancialmente, **clasificatoria**. A partir de una variable dependiente cualitativa y un conjunto de una o más variables independientes cuantitativas (en caso contrario habrá que recodificar los valores de las variables no cuantitativas), el análisis discriminante permite clasificar a los individuos o casos en alguno de los grupos establecidos por la variable dependiente. Desde esta perspectiva, guarda cierta similitud con el análisis de conglomerados o *cluster*: mientras que éste tiene un carácter estrictamente exploratorio (no se conocen *a priori* los grupos a formar), en el análisis discriminante sí se conoce el grupo de pertenencia; en el análisis de conglomerados el objetivo es obtener grupos homogéneos entre sí y heterogéneos respecto a los demás; por su parte, en el análisis discriminante los grupos ya están constituidos debiendo identificar lo específico de cada uno de ellos para poder así asignar a los individuos o casos a cada uno de ellos. En esta finalidad clasificatoria va implícita, pues, una finalidad **descriptiva** (identificar las variables que más y mejor discriminan y caracterizan a los grupos).
- En segundo lugar, y siempre que se apreciara la efectiva diferencia entre grupos, el análisis discriminante cumple una finalidad **predictiva** al proporcionar procedimientos de clasificación sistemática de nuevas observaciones de origen desconocido en uno de los grupos analizados. Desde esta segunda óptica, a su vez, guarda cierta similitud con el modelo de regresión múltiple: mientras que la ecuación de regresión lineal nos permitía estimar directamente el valor de la variable dependiente; en el análisis discriminante será a partir de las puntuaciones

discriminantes desde las que estimaremos la probabilidad de pertenencia a un grupo, y a partir de dicha probabilidad estimaremos a cuál de ellos pertenece cada unidad o caso.

Nosotros vamos a centrar nuestra exposición, exclusivamente, en la primera de las posibilidades implícitas en esta técnica de análisis. Desde esta óptica, el análisis discriminante clasifica a los individuos entre los grupos considerando las variables que más y mejor caracterizan y, en consecuencia, diferencian a los grupos. A estas variables, en el análisis discriminante, se les denominan **variables canónicas o discriminantes**, se presentan como combinaciones lineales de las variables originales y se expresan por una **función discriminante**. Estadísticamente el análisis discriminante busca la obtención de dichas funciones de la siguiente manera:

$$f = g(X_1, X_2, \dots, X_n)$$

en donde:

f es la función discriminante;

g (X₁, X₂, ... X_n) son las funciones lineales de las cinco variables o factores discriminantes introducidos en el análisis.

En este capítulo vamos a ver:

1. El proceso de selección de aquellas variables que, de todas las posibles, más y mejor discriminan a los grupos. El análisis discriminante, al expresar la función discriminante como una combinación lineal, guarda cierta similitud con el análisis de regresión múltiple. De ahí que el procedimiento que exponemos para seleccionar las variables canónicas, el de “paso a paso” o *stepwise*, sea el aplicado en la selección de variables explicativas del modelo de regresión.
2. Una vez seleccionadas las variables discriminantes habrá que valorar la significación estadística de las funciones discriminantes obtenidas. Para ello, utilizaremos el estadístico *Lambda de Wilks*.
3. Si los datos se ajustan al modelo, podremos determinar la función discriminante y, con ella, clasificar.
4. Concluimos el capítulo con la exposición de los pasos y secuencias a seguir para la obtención de los análisis relacionados si es el SPSS el paquete estadístico utilizado.

2. Elección de las variables independientes discriminantes: método “stepwise” o “paso a paso”

En el análisis discriminante el proceso de selección de las variables independientes que más discriminan los grupos establecidos por la variable dependiente se realiza seleccionándolas una a una, “paso a paso”. Este procedimiento implica que: (1) sólo se introducen aquellas variables que más separan a los grupos; (2) sólo se introducen estas variables si cumplen un criterio de entrada; y (3), se eliminan aquellas variables que, previamente seleccionadas, cumplen un criterio de salida. El proceso se inicia con aquella variable que más distancia a los grupos (siempre y cuando supere el criterio de entrada). El proceso finaliza cuando no hay ninguna variable que supere el criterio de entrada; cuando no hay ninguna variable que verifique el de salida; o cuando se haya alcanzado el número máximo de pasos.

1.- Criterio Lambda de Wilks.

El **estadístico Lambda de Wilks** mide las desviaciones que se producen dentro de cada grupo respecto a las desviaciones totales sin distinción de grupos. Si su valor es pequeño, próximo a **0**, la variabilidad total será debida a las diferencias entre grupos y, con ello, las variables con un Lambda de Wilks pequeño serán las que más diferencian, discriminan, a los grupos. Si por el contrario, su valor se aproxima a **1**, los grupos estarán mezclados y, por ello, las variables independientes con un Lambda grande carecen de capacidad discriminante. La primera variable que entrará en el modelo será aquella que de todas presente el Lambda de Wilks más pequeño.

A su vez, el estadístico Lambda de Wilks nos permite contrastar la H_0 de que los centros de los grupos son iguales y, en consecuencia, no existe diferencias entre los mismos. Si el *p-valor* asociado al estadístico es inferior al nivel de significación (normalmente 0.05) rechazamos la hipótesis de igualdad entre los grupos, pudiendo concluir que la información aportada por las respectivas variables es, estadísticamente, significativa.

2.- Verificación del estadístico F de entrada.

Sin embargo, el hecho de que el Lambda de Wilks tome el mínimo valor en una de las variables no implica que éste sea pequeño. De ahí la necesidad de valorar la significación de esta variable. Si al introducir la variable candidata al conjunto de variables independientes se produce un descenso de Lambda de

Wilks, podemos concluir en que la variable es la adecuada. El **estadístico F de entrada** evalúa esa disminución: si la F de entrada asociada a la variable candidata a ser seleccionada es mayor que un determinado valor crítico (normalmente 3.84) se seleccionará.

El **criterio de tolerancia** puede ser aplicado como un criterio adicional a la probabilidad de entrada. Éste nos ayuda a identificar si alguna de las variables del modelo es una combinación lineal de las restantes. Si dicho valor es próximo a 0, la variable analizada será una combinación lineal de las restantes variables independientes introducidas. En síntesis, si la tolerancia para una variable es muy pequeña se excluirá del modelo.

2.- Verificación del estadístico F de salida.

Del mismo modo, y antes de seguir incluyendo variables, debemos comprobar que las variables superan también los criterios de salida. Eliminaremos, definitivamente, aquellas variables candidatas en los pasos correspondientes que al ser suprimidas del conjunto de variables independientes generen un mínimo incremento en el estadístico Lambda de Wilks. El **estadístico F de salida** evalúa el incremento que se produciría en la Lambda de Wilks si la variable candidata fuera eliminada: si la F de salida asociada a la variable candidata a ser eliminada es menor que un determinado valor crítico (normalmente 2.71) se eliminará.

3.- Límite al número de pasos.

Por último, y para evitar que el proceso de selección se convierta en un proceso cíclico se debe establecer un número límite de pasos. Normalmente este límite es el que equivale al doble del número de variables independientes.

Cabe hacer notar que este proceso de selección de variables discriminantes puede ser omitido cuando se tiene la certeza de que las variables con las que se está trabajando son aquellas que: más y mejor describen y discriminan a los grupos; son variables incorrelacionadas; y son variables cuantitativas próximas a una distribución normal. Este es el caso, por ejemplo, de utilizar como variables discriminantes las puntuaciones factoriales obtenidas en análisis previos. En la bibliografía comentada se recoge un estudio en el que se suprime el proceso de selección pues cuenta con las puntuaciones de los factores implícitos o estructurales de un determinado conjunto de variables.

Una vez seleccionadas las variables, ya sea aplicando el método *stepwise* o introduciéndolas todas juntas (aquí solo se exige una tolerancia superior a (1), podemos pasar a la obtención de las funciones discriminantes. Nuestro ejemplo concluye con una única función discriminante. Ésta puede ser consultada en la tabla que en la sección de resultados hemos nombrado como Coeficientes estandarizados de las funciones discriminantes canónicas. Sin embargo, el hecho de que podamos determinarlas no implica que éstas sean significativas.

3. Significación estadística de las funciones discriminantes

Para valorar la significación estadística del conjunto de funciones discriminantes obtenidas volvemos a utilizar el **estadístico Lambda de Wilks**. En este caso, lo debemos consultar en la tabla Resumen de las funciones canónicas discriminantes. En ella se recoge tanto la información relativa a cada una de las funciones como al conjunto de ellas. Las funciones discriminantes generadas serán las mismas en número que las variables seleccionadas. Los estadísticos que se recogen en esta tabla y que nos ayudan a valorar la bondad y representatividad de las funciones son:

1.- Valores del estadístico Lambda de Wilks.

Como ya hemos apuntado este estadístico nos indica cuán diferenciados se encuentran los grupos. Dado que el análisis discriminante va introduciendo cada una de las variables independientes en el modelo según su importancia discriminatoria, las primeras funciones siempre aportarán más información que las últimas.

2.- Significación del estadístico.

Si el *p-valor* asociado al estadístico es inferior al nivel de significación (normalmente 0.05) rechazamos la hipótesis de igualdad entre los grupos, pudiendo concluir que la información aportada por las respectivas funciones discriminantes es, desde un punto de vista estadístico, significativa.

3.- Correlación canónica y autovalor.

Estos dos indicadores son dos medidas que, vinculados con el estadístico Lambda de Wilks, nos permiten conocer qué parte de la información aportada por el conjunto de funciones discriminantes es atribuible a cada una de ellas. Mientras que el primero mide las desviaciones de las puntuaciones discriminantes

entre los grupos respecto a las desviaciones totales sin distinguir grupos; el segundo, mide las desviaciones de las puntuaciones discriminantes entre los grupos. Si el valor de ambos es grande (en concreto para la correlación canónica debe aproximarse a 1), la dispersión viene explicada por las diferencias entre los grupos, discriminando mucho los grupos las funciones correspondientes.

En aquellos ejemplos en los que se el resultado ofrezca más de una función discriminante se puede apreciar como los valores de las columnas de “autovalores” y “correlación canónica” decrecen desde la primera hasta la última. Es la primera función la que alcanza los valores más altos asociándose a ésta el mayor **porcentaje de variabilidad total explicada** (columna porcentaje de la varianza). Normalmente son las primeras funciones las que explican la clasificación obtenida pues son ellas las que explican la casi totalidad de varianza total acumulada (columna “porcentaje acumulado”). Como ya hemos dicho éste no es el caso que se infiere de nuestro ejemplo pues en él sólo se aprecia una única función discriminante que explica el 100% de la variabilidad total.

4. Clasificación de los individuos o unidades

Una vez corroborado la significación estadística de las funciones discriminantes obtenidas e identificado qué parte de la información es atribuible a cada una de ellas, nos resta clasificar a los individuos o unidades de análisis.

Con el fin de obtener la mayor homogeneidad posible se utilizó como criterio de discriminación el valor que alcanzara la función discriminante en cada uno de los individuos. A partir de las funciones discriminantes se obtiene una *regla que permite clasificar* a cada individuo o unidad en uno de los grupos establecidos previamente. Esta regla se basa en las puntuaciones discriminantes de las funciones y en la *probabilidad que a priori* tiene cada unidad de pertenecer a un grupo determinado. Puesto que desconocemos las **probabilidades a priori** consideraremos que éstas son iguales a la proporción de casos en cada uno de los grupos (ver tabla probabilidades previas para los grupos).

A partir de esta regla de clasificación, o bien se ratifica la permanencia de los individuos o unidades en los grupos, o bien, el análisis discriminante en la tabla de Estadísticos de clasificación

por casos nos indica junto a aquellos individuos incorrectamente clasificados su adscripción al grupo más acorde con las funciones discriminantes. Además a partir de esta tabla es posible determinar el **porcentaje de casos correctamente clasificados** siendo éste un índice que nos indica la efectividad de las funciones discriminantes y que siempre hay que valorar en relación a los porcentajes de casos clasificados correctamente *a priori*.

En la tabla de resultados de clasificación se muestra en la parte superior los resultados considerando el número de casos y en la parte inferior expresados en porcentajes. En nuestro caso el porcentaje de casos correctamente clasificados asciende a un total de 67.4%, lo que nos permite ratificar la significación estadística de la función discriminantes.

5. Cuadro de Diálogo del Análisis Discriminante

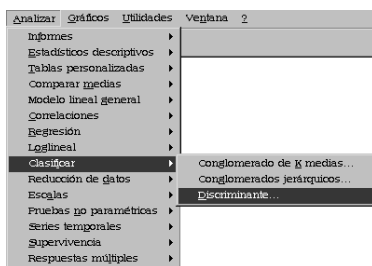


Figura 1

1º paso: Para acceder al cuadro principal del análisis discriminante y sabiendo que ésta es una técnica clasificatoria, seleccionamos en el menú **Analizar: Clasificar: Discriminante** (figura 1).

2º paso: Del listado de variable que aparece en la ventana de la izquierda seleccionamos la **Variable de agrupación**, o lo que es lo mismo, la variable dependiente a partir de la cual vamos a agrupar a los individuos en función a la respuesta ofrecida en cada una de sus categorías (figura 2).

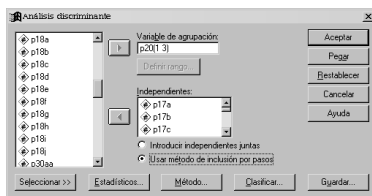


Figura 2

En nuestro ejemplo, la variable de agrupación es la P20. En ellas se le pregunta al encuestado que opinara sobre la política que se debería adoptar hacia los inmigrantes de países menos desarrollados. Las dos posibles respuestas eran: favorecer su integración; o, por el contrario, favorecer su regreso. De esta forma, agruparemos a los individuos en una política u otra (no se consideran los NS/NC de la pregunta)

3º paso: A continuación seleccionamos las variables independientes que vamos a utilizar en nuestro análisis colocándolas en la ventana de **Independientes** (figura 2). Las variables que vamos a utilizar en nuestro ejemplo son las P17 a, b, c,..., j (quedan excluidos del análisis los NS/NC). En ella se pedía a los encuestados que se posicionaran en una escala de 0 (muy mal)

a 10 (muy bien) sobre sus simpatías respecto a grupos de distintas nacionalidades.

El propósito que se persigue al aplicar el análisis discriminante sobre el conjunto de variables (dependiente e independientes) seleccionadas es el de comprobar, a partir del conjunto de las variables independientes si un individuo que en la variable P20 había sido clasificado en un grupo se encuentra correctamente clasificado o de otro modo, ese individuo debería pertenecer a otro grupo. Sabiendo si una persona tienen o no simpatías hacia unos determinados inmigrantes, podremos saber si está bien clasificado en una de las políticas de inmigración que se presentan en la variable P20. Lo lógico, es que si una persona tiene gran simpatía hacia los inmigrantes africanos, pertenezca al grupo de individuos que están de acuerdo con la política de inmigración tendente a favorecer la integración de este colectivo.

Antes de llegar a estas conclusiones debemos encontrar del conjunto de las variables independientes aquellas que desempeñan un papel discriminante; esto es, aquellas que nos van a permitir demostrar si una persona en función de la variable dependiente está bien agrupada.

4º paso: Las variables independientes se pueden introducir juntas o bien se pueden ir introduciendo una o una. Ambas opciones, **Introducir independientes juntas** y **Usar método de inclusión por pasos**, aparecen en el cuadro de diálogo principal del análisis discriminante. Con la primera opción se introducen de forma simultánea aquellas variables que satisfacen el criterio de tolerancia. Ésta es una opción muy forzada o burda de ahí que se aconseje aplicar la segunda de las opciones. Si es ésta la opción elegida, deberemos especificar qué método es el que vamos a seguir en el procedimiento de paso a paso.

5º paso: Al seleccionar la segunda opción automáticamente se activa el botón de comando **Método de inclusión por pasos** situado en la parte inferior del cuadro de diálogo principal. Al acceder a este subcuadro de diálogo deberemos especificar: que el **Método** que vamos a aplicar es el de **Lambda de Wilks**; y que los **Criterios** de verificación son los asociados al **valor de F**. Por defecto los valores críticos que aparecen son: de 2.71, para el de **Salida**; y 3.84 para el de **Entrada**.

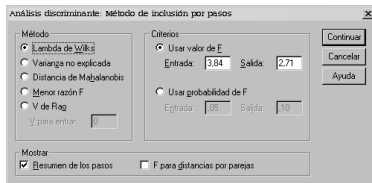


Figura 3

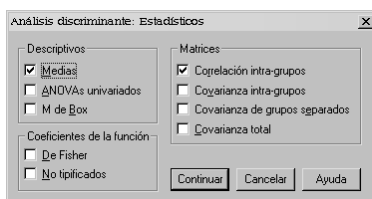


Figura 4

Por último, y todavía en el subcuadro de diálogo de Método, nos interesa que en las tablas de resultados aparezca un **Resumen de los pasos**. En ella aparece el valor del estadístico de lambda para cada uno de los pasos, así como su significación (figura 3).

6º paso: Cliqueando en el botón de comando de **Estadísticos**, situado en la parte inferior del cuadro de diálogo principal del análisis discriminante, accedemos al subcuadro de diálogo que nos permite seleccionar dentro de los **Descriptivos** a las **Medias** y dentro de las **Matrices** la de **Correlación intra-grupos** (figura 4).

Una vez que tengamos seleccionadas las variables que entrarán en el modelo, pasamos a determinar si los individuos en función de estas variables, están bien o incorrectamente clasificados. Para ello resulta imprescindible poder comparar los resultados que obtengamos al aplicar el análisis discriminante, la clasificación, con las probabilidades previas (a priori) de pertenencia a los grupos. A su vez, ya hemos comentado que el porcentaje de casos correctamente clasificados es un buen índice de la efectividad de las funciones discriminantes. Éste siempre habrá que valorarlo en relación a los porcentajes de casos clasificados correctamente *a priori*.

7º paso: Clicando sobre el botón de comando **Clasificación** podemos especificar bajo qué términos considerar a las **Probabilidades previas** pues, lo más frecuente, es que desconozcamos dichos valores. Dos son las posibilidades:

- Podemos considerar que las probabilidades a priori de pertenecer a los grupos es para todos la misma eligiendo la opción **Todos los grupos iguales**.
- O bien, podemos considerar que las probabilidades previas equivalen a la proporción de casos que hay en cada grupo. Esta es una opción que se ajusta más a la realidad de nuestros datos y por ello hemos seleccionado la opción **Calcular según tamaños de grupos** (figura 5).

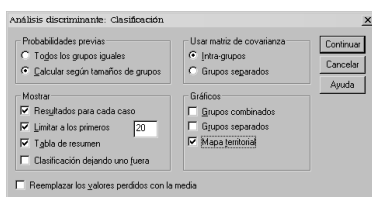


Figura 5

Por lo que respecta al proceso de clasificación escogido, nos interesa que aparezcan los resultados del mismo. En **Mostrar**, solicitamos:

- Los **Resultados para cada caso** aunque **limitando la presentación** a los primeros **20** casos.
- Resulta también de gran ayuda la consulta de la **Tabla de resumen** o **matriz de confusión**. En ella se indica: el grupo real de pertenencia y el grupo predicho; y el número y porcentaje de casos correctamente clasificados.

Para terminar, y dentro de esta subventana de clasificación, le pedimos al programa que elabore el **Mapa territorial** de los dos grupos de nuestro ejemplo. Los números representan las fronteras entre grupos y los asteriscos el centroide del grupo.

5. Bibliografía Comentada

- Flavián, Carlos et al. (1994): “Factores determinantes del nivel de equipamiento doméstico. Un contraste empírico”, *ESIC*, n° 85.

A partir de la constatación de que en los últimos años se ha producido un importante incremento del equipamiento doméstico en los hogares españoles, esta investigación plantea analizar si esta circunstancia está relacionada con hogares de mayor nivel socioeconómico. En este estudio se consideran productos de consumo que facilitan las tareas domésticas o bien productos que dadas sus particularidades se asocian a familias con menor disponibilidad de tiempo. Por su parte, y como variables que miden el nivel socioeconómico se contemplan los ingresos, el nivel de estudios de la pareja, el tamaño de la unidad familiar, la profesión desempeñada y el número de automóviles. Como técnica de análisis se ha aplicado el discriminante la cual incluye a los entrevistados en dos grupos, a saber: adoptan o no cada tipo de producto. El análisis concluye en que no se puede rechazar la hipótesis nula de que existe una adopción entre adopción de equipamiento doméstico y el nivel socioeconómico de las unidades familiares. Esta afirmación se basa en la presencia de variables estadísticamente significativas en la mayoría de las funciones de clasificación estimadas, así como en el porcentaje de correcta clasificación (superior al 63%).

- Martínez Ramos, Emilio (1984): “Fundamentos del análisis discriminante y su aplicación en un estudio electoral”, en J.J. Sánchez Carrión (ed.) (1984), *Introducción a las técnicas de análisis multivariable aplicadas a las ciencias sociales*, Madrid, Centro de Investigaciones Sociológicas (CIS), pp. 139-169.

En este capítulo el autor de una forma muy didáctica, clara y sencilla explica cada uno de los pasos o secuencias a tener en cuenta cuando es el análisis discriminante la técnica aplicada. El ejemplo que utiliza para tal fin es una encuesta que sobre intención de voto (elecciones legislativas de 1982) declarada por 492 individuos.

6. Resultados

Los resultados obtenidos al aplicar la técnica discriminante son los que enunciamos a continuación. Nótese que éstos responden a las restricciones y peticiones impuestas al programa y que, en consecuencia, los resultados variarán en función de los intereses del propio investigador.

- En primer lugar, y tal y como viene siendo costumbre en todos los análisis expuestos, la primera salida recoge los **estadísticos básicos** de las variables seleccionadas así como un **resumen del proceso**.
- A continuación, y como previo paso a la concreción de las funciones discriminantes, se presenta la **tabla resumen de variables discriminantes seleccionadas**. La selección de aquellas variables que más y mejor discriminan a los grupos sigue el proceso de paso a paso. Le sigue la tabla de **variables que entran en el modelo** y la tabla de **variables excluidas del modelo**.
- Una vez relacionados los estadísticos que avalan la significación estadística de las variables incluidas, el programa nos ofrece un **resumen de las funciones canónicas discriminantes**. Esta tabla incluye medidas (lambda de Wilks) que nos permiten valorar no solo la significación estadística de las funciones discriminantes obtenidas (en nuestro caso de una única función discriminante) sino también saber qué parte de la información aportada por

el conjunto de funciones discriminantes es atribuible a cada una de ellas (en nuestro caso y teniendo en cuenta que solo tenemos una función discriminante el 100% de la variabilidad total es explicado por dicha función). Los **coeficientes de la función discriminante** seleccionada aparecen a continuación.

- Concluido el proceso de elección de las funciones discriminantes el programa pasa a clasificar los individuos en los grupos establecidos. A partir de, en nuestro caso, la función discriminante, se obtiene una regla que permite clasificar cada una de las unidades de análisis en los grupos establecidos previamente. Esta regla se basa en las puntuaciones discriminantes de la función seleccionada y en la probabilidad que a priori tiene cada individuo de pertenecer a un grupo determinado. Por ello a las tablas comentadas le sigue la tabla de **probabilidades previas para los grupos** y la tabla de **estadísticos de clasificación** por casos. En esta última aparecen los casos incorrectamente clasificados indicando el mejor grupo de pertenencia.
- Finaliza el análisis con una tabla **resumen de los resultados de clasificación**. En la parte superior aparecen los resultados considerando el número de casos y en la inferior expresados en porcentajes. Además ofrece el porcentaje de casos correctamente clasificados, índice de efectividad de la función discriminante.

6.1. Resumen del Procesamiento de los Datos

Casos no ponderados		N	Porcentaje
Válidos		949	79,1
Excluidos	Código de grupo de perdido o fuera de rango	97	8,1
	Perdida al menos una variable discriminante	145	12,1
	Perdidos o fuera de rango ambos, el código de grupo y al menos una de las variables discriminantes.	9	,8
	Total	251	20,9
Total		1200	100,0

6.2. Estadísticos Básicos

En relación con los inmigrantes de países		N válido (según lista)			
		Media	Desv. típ.	No ponderados	Ponderados
Favorecer su integración	Asiáticos orientales	6,67	2,27	539	539,000
	Europeos de este norteamericanos	6,80	2,25	539	539,000
	Rusos	5,92	2,58	539	539,000
	arabes	6,42	2,28	539	539,000
	Europeos occidentales	5,94	2,64	539	539,000
	Gitanos	6,70	2,12	539	539,000
	judíos	5,61	2,66	539	539,000
	sudamericanos	6,26	2,49	539	539,000
	Africanos de raza negra	6,60	2,23	539	539,000
Favorecer su regreso	Asiáticos orientales	6,42	2,31	539	539,000
	Europeos de este norteamericanos	5,66	2,33	410	410,000
	Rusos	6,00	2,38	410	410,000
	arabes	5,50	2,50	410	410,000
	Europeos occidentales	5,53	2,47	410	410,000
	Gitanos	4,80	2,70	410	410,000
	judíos	6,12	2,15	410	410,000
	sudamericanos	3,85	2,71	410	410,000
	Africanos de raza negra	5,07	2,71	410	410,000
Total	Asiáticos orientales	5,57	2,34	410	410,000
	Europeos de este norteamericanos	5,00	2,50	410	410,000
	Rusos	6,23	2,35	949	949,000
	arabes	6,45	2,34	949	949,000
	Europeos occidentales	5,74	2,55	949	949,000
	Gitanos	6,04	2,40	949	949,000
	judíos	5,45	2,72	949	949,000
	sudamericanos	6,45	2,15	949	949,000
	Africanos de raza negra	4,85	2,82	949	949,000

6.3. Elección de las variables independientes discriminantes. Resumen de método STEPWISE o paso a paso

Variables introducidas/eliminadas^{a,b,c,d}

Paso	Introducidas	Estadístico	gl1	gl2	gl3	Lambda de Wilks			
						F exacta			
						Estadístico	gl1	gl2	Sig.
1	Gitanos	,904	1	1	947,000	100,436	1	947,000	,000
2	Africanos de raza negra	,892	2	1	947,000	57,174	2	946,000	,000
3	norteamericanos	,879	3	1	947,000	43,439	3	945,000	,000
4	Asiáticos orientales	,873	4	1	947,000	34,283	4	944,000	,000

En cada paso se introduce la variable que minimiza la lambda de Wilks global.

^a. El número máximo de pasos es 20.

^b. La F parcial mínima para entrar es 3,84.

^c. Maximum partial F to remove is 2,71.

^d. El nivel de F, la tolerancia o el VIN son insuficientes para continuar los cálculos.

6.4. Variables que entran en el modelo siguiendo el proceso “paso a paso”

Variables en el análisis

Paso		Tolerancia	F que eliminar	Lambda de Wilks
1	Gitanos	1,000	100,436	
2	Gitanos	,631	29,745	,920
	Africanos de raza negra	,631	12,675	,904
3	Gitanos	,590	38,923	,915
	Africanos de raza negra	,571	20,705	,898
	norteamericanos	,691	14,353	,892
4	Gitanos	,589	37,169	,908
	Africanos de raza negra	,501	11,322	,884
	norteamericanos	,606	19,516	,891
	Asiáticos orientales	,577	6,112	,879

6.5. Variables que no entran en el modelo siguiendo el proceso “paso a paso”

Variables no incluidas en el análisis

Paso		Tolerancia	Tolerancia mín.	F que introducir	Lambda de Wilks
0	Asiáticos orientales	1,000	1,000	44,751	,955
	Europeos de este	1,000	1,000	28,085	,971
	norteamericanos	1,000	1,000	6,461	,993
	Rusos	1,000	1,000	33,038	,966
	árabes	1,000	1,000	42,247	,957
	Europeos occidentales	1,000	1,000	16,987	,982
	Gitanos	1,000	1,000	100,436	,904
	judíos	1,000	1,000	49,351	,950
	sudamericanos	1,000	1,000	47,501	,952
	Africanos de raza negra	1,000	1,000	82,112	,920
1	Asiáticos orientales	,810	,810	6,017	,898
	Europeos de este	,846	,846	1,986	,902
	norteamericanos	,764	,764	6,369	,898
	Rusos	,759	,759	,810	,903
	árabes	,683	,683	,968	,903
	Europeos occidentales	,798	,798	,168	,904
	judíos	,644	,644	1,524	,903
	sudamericanos	,740	,740	3,877	,900
	Africanos de raza negra	,631	,631	12,675	,892

(continúa...)

Variables no incluidas en el análisis

Paso		Tolerancia	Tolerancia mín.	F que introducir	Lambda de Wilks
2	Asiaticos orientales	,658	,513	1,007	,891
	Europeos de este	,695	,519	,011	,892
	norteamericanos	,691	,571	14,353	,879
	Rusos	,653	,543	,210	,892
	arabes	,598	,553	,082	,892
	Europeos occidentales	,694	,549	3,259	,889
	judios	,520	,510	,127	,892
	sudamericanos	,553	,472	,043	,892
3	Asiaticos orientales	,577	,501	6,112	,873
	Europeos de este	,575	,510	2,578	,876
	Rusos	,493	,493	2,616	,876
	arabes	,520	,520	1,325	,878
	Europeos occidentales	,533	,531	,000	,879
	judios	,503	,488	,116	,879
	sudamericanos	,514	,461	1,586	,877
4	Europeos de este	,382	,382	,046	,873
	Rusos	,455	,455	,935	,872
	arabes	,442	,442	,043	,873
	Europeos occidentales	,485	,485	,579	,873
	judios	,459	,459	,165	,873
	sudamericanos	,474	,440	,354	,873

6.6. Valores del estadístico Lambda de Wilks y su significación

Lambda de Wilks									
Paso	Número de variables	Lambda	gl1	gl2	gl3	F exacta			
						Estadístico	gl1	gl2	Sig.
1	1	,904	1	1	947	100,436	1	947,000	1,846E-22
2	2	,892	2	1	947	57,174	2	946,000	,000
3	3	,879	3	1	947	43,439	3	945,000	,000
4	4	,873	4	1	947	34,283	4	944,000	,000

6.7. Resumen de las funciones canónicas discriminantes

Autovalores

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	,145 ^a	100,0	100,0	,356

^a. Se han empleado las 1 primeras funciones discriminantes canónicas en el análisis.

Lambda de Wilks

Contraste de las funciones	Lambda de Wilks	Chi-cuadrado	gl	Sig.
1	,873	128,178	4	,000

6.8. Coeficientes estandarizados de las funciones discriminantes canónica

	Función
	1
Asiáticos orientales	,296
norteamericanos	-,513
Gitanos	,712
Africanos de raza negra	,432

Matriz de estructura

	Función
	1
Gitanos	,854
Africanos de raza negra	,773
judios ^a	,624
Asiáticos orientales	,570
sudamericanos ^a	,550
arabes ^a	,542
Europeos de este ^a	,440
Rusos ^a	,430
Europeos occidentales ^a	,400
norteamericanos	,217

Correlaciones intra-grupo combinadas entre las variables discriminantes y las funciones discriminantes canónicas tipificadas
Variables ordenadas por el tamaño de la correlación con la función.

^a. Esta variable no se emplea en el análisis.

Probabilidades previas para los grupos

En relación con los inmigrantes de países menos	Casos utilizados en el análisis		
	Previas	No ponderados	Ponderados
Favorecer su integración	,568	539	539,000
Favorecer su regreso	,432	410	410,000
Total	1,000	949	949,000

6.9. Estadísticos de clasificación por casos

Estadísticos por casos												
Original	Número de caso	Grupo mayor					Segundo grupo mayor					Puntuación discriminante antes
		Grupo real	Grupo pronosticado	P(D>d G=g)		Distancia de Mahalanobis al cuadrado hasta el centroide	Grupo	P(G=g D=d)	Distancia de Mahalanobis al cuadrado hasta el centroide	Función 1		
				p	gl							
1	desagrupado	1	,656	1	,556	,198	3	,444	,105	-,113		
2	1	1	,656	1	,556	,198	3	,444	,105	-,113		
3	1	1	,155	1	,841	2,023	3	,159	4,800	1,754		
4	desagrupado	3	,841	1	,544	,040	1	,456	,940	-,638		
5	3	1**	,656	1	,556	,198	3	,444	,105	-,113		
6	1	1	,477	1	,506	,505	3	,494	,003	-,379		
7	3	3	,271	1	,704	1,211	1	,296	3,494	-1,537		
8	1	1	,449	1	,760	,573	3	,240	2,327	1,089		
9	1	1	,155	1	,841	2,023	3	,159	4,800	1,754		
10	3	3	,744	1	,568	,107	1	,432	1,200	-,763		
11	1	1	,763	1	,690	,091	3	,310	1,146	,634		
12	3	1**	,477	1	,506	,505	3	,494	,003	-,379		
13	1	1	,763	1	,690	,091	3	,310	1,146	,634		
14	1	1	,655	1	,714	,200	3	,286	1,478	,779		
15	1	1	,907	1	,659	,014	3	,341	,785	,450		
16	1	1	,943	1	,626	,005	3	,374	,486	,260		
17	1	1	,626	1	,548	,238	3	,452	,079	-,156		
18	3	3	,936	1	,521	,006	1	,479	,720	-,517		
20	3	3	,375	1	,669	,787	1	,331	2,741	-1,324		
21	3	1**	,735	1	,577	,115	3	,423	,185	-,007		

**. Caso mal clasificado

** Caso mal clasificado

6.10. Resumen de los resultados de clasificación

Resultados de la clasificación^a

		En relación con los inmigrantes de países menos desarrollados	Grupo de pertenencia pronosticado		Total
			Favorecer su integración	Favorecer su regreso	
Original	Recuento	Favorecer su integración	475	99	574
		Favorecer su regreso	230	204	434
		Casos desagrupados	73	27	100
	%	Favorecer su integración	82,8	17,2	100,0
		Favorecer su regreso	53,0	47,0	100,0
		Casos desagrupados	73,0	27,0	100,0

^a Clasificados correctamente el 67,4% de los casos agrupados originales.